



# Pipelines de ETL e Machine Learning com Apache Spark

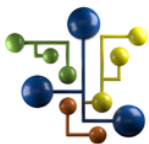
## Projeto 3 Pipeline de Machine Learning em Cluster Spark Para Previsão de Churn Treinamento e Deploy



Neste projeto vamos construir um pipeline de Machine Learning para prever o Churn com algoritmo RandomForest. O modelo vai ler os dados a partir do HDFS, executar o processamento e treinamento com Spark MLlib e salvar o resultado de volta no HDFS, tudo sendo executado em um ambiente distribuído, com múltiplas máquinas em um cluster. Vamos ainda salvar o modelo treinado no armazenamento distribuído.

Com o modelo treinado, avaliado e salvo no HDFS, faremos o deploy usando o modelo para previsões com novos dados.

Bons estudos.



**Equipe DSA**

Muito Obrigado!  
Continue Trilhando Uma Excelente Jornada de Aprendizagem.